Parametric Timing Failures and Defect-based Testing in Nanotechnology CMOS Digital ICs

Chuck Hawkins, Ali Keshavarzi*, Jaume Segura**

Univ. New Mexico, Intel Corp.*, Univ. Balearic Islands**

hawkins@eece.unm.edu

Abstract

Parametric failures have been with us since the beginning of CMOS technology, but their significance is now more serious and growing. We refer to ac parametric timing failures that fall into two classes: (1) intrinsic ICs (free of defects), and (2) extrinsic ICs (presence of defects. The succession of 180 nm, 130 nm, and 90 nm CMOS IC technologies show an increasing lack of manufacturing control of circuit parameter variance. The spread of statistical parameters from their target, such as L_{eff}, V_t, metal line width and spacing, F_{MAX}, $V_{DD}(min)$, etc., presents subtle failures or degradations called parametric (timing) failures. Our presentation will show fab statistics of these parameters and how they cause peculiar failures. The properties of intrinsic and extrinsic ICs are summarized in a defect-based test (DBT) analysis showing what detection approaches are affected and those that aren't.

I. Introduction

Parametric failures have unusual properties that form broad behavioral patterns. Parametric timing failures fall into two classes: (1) intrinsic ICs (free of defects), and (2) extrinsic ICs (presence of defects). Intrinsic ICs can fail due to an unfortunate distribution of circuit statistics. An extrinsic IC may have subtle imperfections in vias or contacts that cause its speed characteristics to worsen at lower temperatures. We will discuss each class with supporting behavioral data, and then discuss the challenge of detection.

II. Intrinsic ICs

Two factors cause intrinsic parameter variation: environmental and physical. Environmental factors include variation of the power supply levels within the die, or on the board, or switching activity and temperature variation across the circuit. Physical variation comes from the inherent weaknesses in IC manufacturing control that allow transistor and interconnect structural variations. These deviations from targeted values are limitations or imperfections in process and mask steps. The random and uncontrollable nature of parameter variations cause these failures to be non-systematic from die-to-die, wafer-to-wafer, and even transistor-to transistor properties within a die (intra-die variation) [Ors02, Seg02]. For example, a drive transistor with a current strength slightly above the nominal value may compensate a slightly high interconnect resistance in a signal path. This same high resistance may lead to an unacceptable signal delay if the transistor has a driving strength below or at the nominal conductance value. This leads to inaccuracies that impact circuit quality, and can provoke erroneous behaviors that occur at very specific circuit states or environmental conditions. Recent technologies show die-to-die reordering of critical paths [Ors02].

These types of failures are difficult to detect and locate. Several failure analysis experiences with parametric delay defects showed that one to three months of effort may be necessary to locate one defect on an IC. This is intolerable, and research efforts are currently directed at more efficient location techniques [Bru02, Col01].

Present design technology is unable to characterize the whole complexity of parameter combinations, so present strategies check only the corner parameters. Therefore, unfortunate parameter combinations can be very difficult to detect and screen. For example, an IC that is normally on the fast edge of the distribution could have a delay defect that puts it now at the slow, but acceptable range and the part passes. Defective parts that pass function are an increased reliability risk [Rig98]. Test limits can no longer rely on single limit approaches, but rather statistical techniques that improve the test signal-to-noise ratio.

Figure 1 shows threshold voltages measured for transistors designed with two different target channel lengths, while the inset illustrates the threshold voltage dependence with the effective transistor length. The spread in effective transistor length is typical of parameter variations in sub-micron

technologies. Ideally, since all devices are designed to have one of the two target transistor lengths, only two points should appear in the graph. In practice, the spread in transistor effective length translates to a spread in device threshold voltages and circuit speed performance.

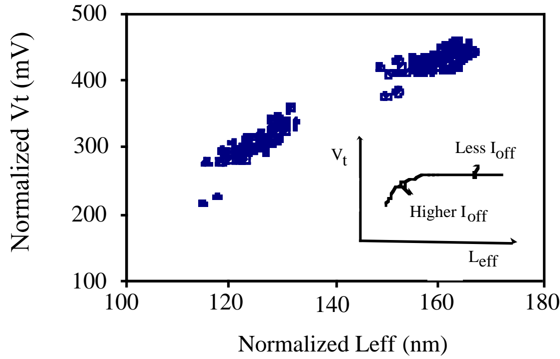


Fig. 1. Measured transistor threshold voltages for two set of *n*MOS devices with different designed channel lengths. The inset plot shows the V_t roll-off dependence with the effective transistor length. (Permission of Intel Corporation)

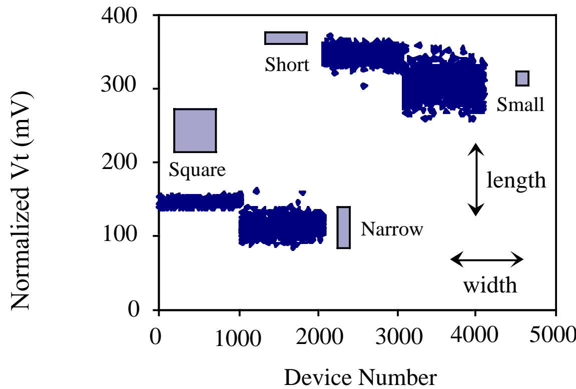


Fig. 2. Measured threshold voltages for four sets of *n*MOS transistors. Square refers to transistors designed with large length and width, while small refers to devices with minimum length and width. Short devices are wide minimum length transistors, while narrow transistors have large L and small W [Seg02].

Transistor threshold voltages can vary with polysilicon gate length and width. Figure 2 shows threshold voltages measured from 4,000 *n*MOS transistors of different sizes in an Intel test chip. These differences in V_t directly affect individual transistor current drive and chip differences in speed. The narrow width transistors (<0.4 μm) show marked increase in variance between devices. Narrow width effects are expected on technologies less than 180 nm where shallow trench isolation tends to decrease the effective threshold voltage and increase V_t variance [Bur98]. Short devices have larger transistor threshold voltages than the square devices due to the drain-source halo implant technology that controls V_t for these devices. The short and small sizes in Fig. 2 are the typical gate dimensions found in normal digitally designed ICs.

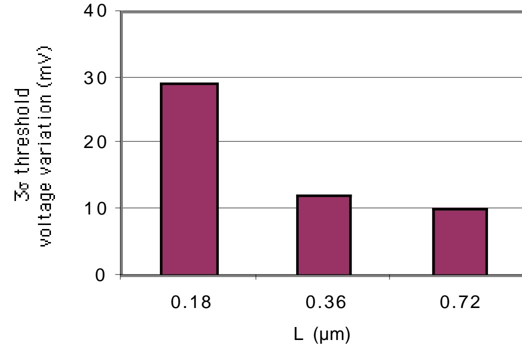


Fig. 3. 3σ variation of transistor threshold voltage at three technology nodes [Nar99].

Figure 3 shows the 3σ variation of transistor threshold voltage at three technology nodes. At the 180 nm, the variation in V_t is markedly increased. This occurs while the absolute value of threshold is reduced from about 0.45-0.4 V at the 0.36 μm and 0.18 μm nodes to about 0.3 V at the 130 nm node [21]. Transistor threshold voltages can vary within a die. Figure 4 shows variation from 8% at the 180 nm die to more than 12% for the 130 nm die [Nar99]. Keshavarzi, et. al., also presented L_{eff} variation data and its impact on I_{Dsat} which is the transistor speed parameter [Kes97]. Transistor L_{eff} intrinsic variation is the primary variable that influences F_{max} [Bur98].

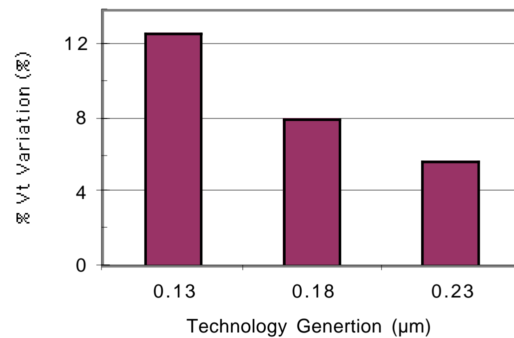


Fig. 4. Threshold voltage variation within the die [Nar99].

Defect-free ICs also show increased random failures due to interconnect properties [Ber98]. These properties include crosstalk influence on timing or even Boolean failure [Bre96]. Crosstalk errors arise from poor design rule implementation (a difficult problem), or statistical fluctuations in metal line spacing (and metal width) [Deu01, Syl01]]. These properties are aggravated by the present geometric structure of metals in which their height to width ratios (aspect ration) are now >1. Metal interconnect aspect ratios of 2-3 make the dominant metal capacitance from sidewall to sidewall instead of to the power rails as in older technologies.

Although crosstalk and switching noise are traditionally categorized as design–related problems, technology-scaling complexity has diffused the boundaries between design and functionality, so that designers must take statistical variation into account in the designs, so that some circuits from the same fabrication lot fail, while others do not.

Another intrinsic concern is the rate of change in power supply currents that is now on the order of tens of Amps per ns. Low V_{DD}'s with higher di/dt increase the statistical risk of parametric failures due to inductive ground bounce and power supply IR drop, and to lower signal margin.

Each technology node has a minimum metal width of about that technology node number. A 100 nm node has a minimum metal width of about 100 nm [ITR01]. The minimum dielectric spacing between minimum metal geometry is also about 100 nm. The vias and contacts may have diameters on the order of the metal widths, and since the intermetal dielectric vertical spacing scales more slowly, the via and contact aspect ratios tend to get larger. ICs have hundreds of millions and billions of these structures and defect-free vias cannot be guaranteed. Vias and contacts are a difficult challenge for fabrication, test engineering, and failure analysis.

Figure 5 shows the statistical timing variation measured on 25 wafers with cumulative propagation delays measured on 910 die in each wafer [Bak99]. The initial rise in the plots shows a near straight line relation to the normal distribution. The difference in propagation delay between fastest and slowest die at V_{DD} = 2.5 V is about 1.7 ns representing a difference of about 24% with respect to the faster IC. When the power supply was reduced to V_{DD} = 1.2 V (Fig. 5b), the propagation delays increased almost three times. The break in the curve at the 95% cumulative point is more distinct. Test limits could be set tightly at the 95% cumulative point to reject the outliers for high reliability products, or set somewhat higher trading yield loss against product reliability expectations, however tight limit setting is not practical.

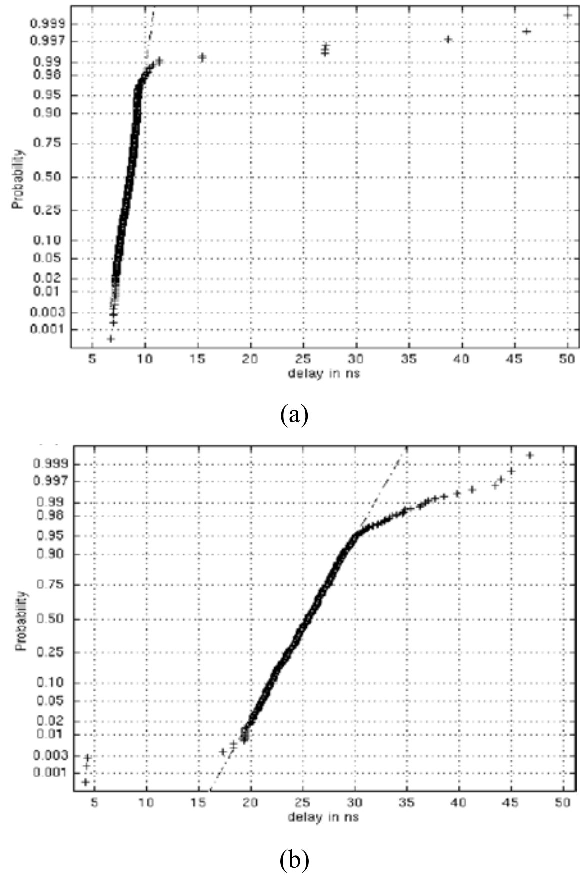


Fig. 5. Cumulative distributions for propagation delay from 25 wafers with data including 910 die per wafer. (a) V_{DD} = 2.5 V, (b) V_{DD} = 1.2 V [Bak99].

There is a physical difference between parts in the normal and outlier distributions. The outliers have defects causing delay in addition to the part's otherwise normal variation. An important point is that real tests that target delay faults do not have the capability and resolution to measure to this fine degree for each delay path. Each 2-vector timing test pattern does not have an individual test limit, so that delay fault testing typically uses the period of the system clock, and adds an amount to account for tester noise and parameter variance. Delay fault testing is a gross test for timing errors.

Similar results were obtained by [Nat98] from process variation data measured on a 0.8 µm process. A test limit must take into account the slowest parameter measurement. Slower outliers from the fastest die overlap normal data in the slower die. This is part of the challenge in detecting these statistical variance failures.

Figure 6 shows the sensitivity of F_{MAX} to V_{DD} of an Intel 1 GHz, 6.6 million transistors, router chip built in 150 nm technology. The chip has an approximate 1.8 MHz change in F_{MAX} per mV of power supply voltage at V_{DD} = 1.3 V. Figure 6 shows about a 14% increase in F_{MAX} when V_{DD} increases 10% from V_{DD}=1.3 V. The sensitivity is higher at the low end V_{DD} and saturates above V_{DD} = 1.6 V. Bernstein, et al., reported a change in F_{MAX} with V_{DD} of 200 kHz/mV for a 180 nm microprocessor, and gave a performance rule of thumb that chips vary by 7-9% when the power supply varies by about 10% [Ber98]. Vangal, et al., showed F_{MAX} versus V_{DD} plots for a 130 nm dual-Vt 5 GHz technology with a sensitivity of 11.3 MHz/mV at V_{DD} = 1.0 V [Van02]. Figure 6 emphasizes that mV noise changes in V_{DD} induced on a critical signal node in normal operation can measurably affect IC speed. A related system problem occurs when board power supplies have a ± 5% accuracy. This could move a test verified 1 GHz IC performance in Fig. 6 to an operational range from about 883 MHz to 1.18 GHz on the board. One protection used by manufacturers is to guardband the shipped product anticipating power supply influences, but this is wasteful of IC capability and also a yield loss.

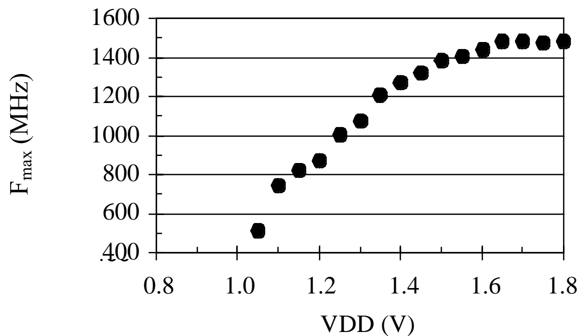


Fig. 6. F_{MAX} versus V_{DD} for a 150 nm IC [Seg02].

Figure 7 shows temperature data taken from a small, 20,000 transistors test circuit [Seg02]. The temperature sensitivity in the regression equations in Figure 7 is –13.1 kHz/°C and –12.3 kHz/°C for the fast and slow ICs. The major effect that slows an IC with temperature rise is the decrease in carrier mobility. A compensating speed effect is that the absolute values of the *n*- and *p*-channel transistor threshold voltages decrease as temperature rises. For V_{DD} = 0.5 V and V_t = 200 mV technologies, the compensation of the threshold behavior can even lead to ICs with a positive temperature performance [Kan01]. There may be thermal instability issues, but temperature characterization must be known for each technology to allow recognition of anomalous parametric behavior.

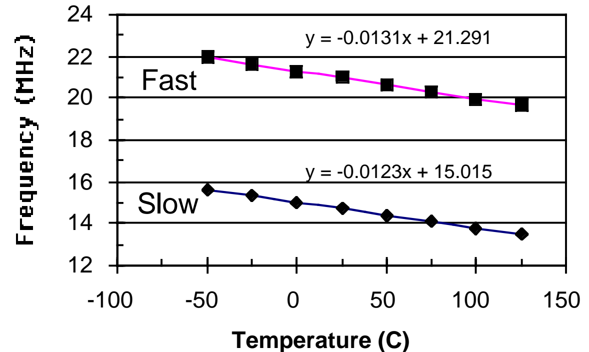


Fig. 7. Speed vs. temperature dependence on a 20k transistor circuit.

III. Extrinsic ICs

We will describe four extrinsic IC mechanisms associated with parametric failures: (1) resistive vias and contacts, (2) metal mousebites, (3) metal slivers, and (4) gate oxide shorts in ultrathin technologies. Resistive vias and contacts are a major defect-related parametric failure mechanism. Mousebites occur when sections of metal are missing from an interconnect line. Slivers are a common defect in which a metal particle lies between two metal conductors and barely contacts the signal lines. The gate oxide short shows a variety of responses. Some gate shorts show timing and power supply dependent failures, and the recent ultrathin oxides have a unique failure mode questioning the implied reliability risk for some gate oxide shorts. Typically the most common extrinsic parametric failures now are the metal vias and contacts, and the presence of metal slivers aggravated by CMP.

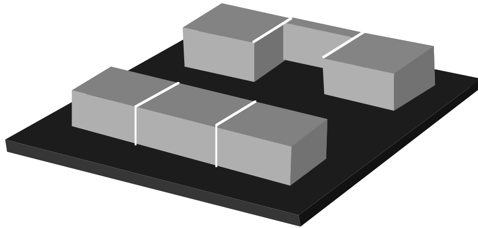
Resistive Vias

Modern ICs can have billions of transistors and perhaps ten times that number of vias. Modern via aspect ratios are now > 5, so that defective vias with elevated resistance are not surprising. Vias and contacts have different sizes depending upon the metal level. The vias and contacts at the lowest metal level are the smallest usually close to minimum feature size. Cracks in flat metal lines also show properties similar to resistive vias, but are a less common failure mechanism, particularly with the shunting barrier metals deposited around the Al or Cu metal interconnect.

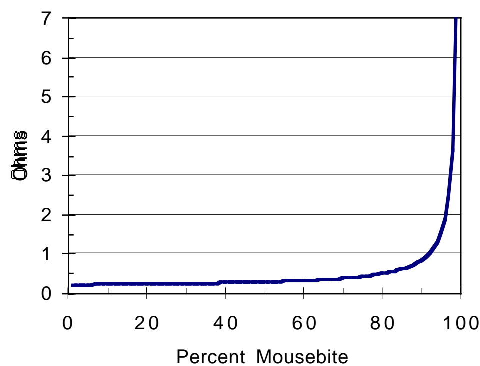
Resistive vias have a unique, but expensive signature in test. A speed parameter (F_{MAX} or prop delay) will show degradation as the temperature is reduced. The problem was severe enough even at 0.25 um technologies to require two temperature testing.

Metal Mousebites

Missing regions of interconnect metal are called mousebites. They can be due to particle defects, electromigration, or stress voids. Mousebites have a minor electrical effect, but are a major reliability risk [Seg02]. Figure 8a sketches a defect-free and a defective (mousebite) section of interconnect. If sheet resistance is $R_{\square} \approx 70$ mΩ/□ and a 90% bite is taken out of the middle section, then that line resistance changes from 210 mΩ to 840 mΩ. This major defect in the line would not elevate resistance sufficiently to be detected by a speed test. Analysis shows that line resistance rises exponentially, but not significantly until the mousebite exceeds about 95% (Figure 8b). These conclusions extend to voided vias and contacts that also show this resistance dependency on volume voiding. The via or contact must be well voided to cause an RC delay failure sensitive to temperature.



(a)



(b)

Fig. 8. (a) Control metal line and (a) normal metal line and one with mousebite, (b) Voided metal resistance (mΩ) versus percent metal voiding using R = 70 mΩ/□ [Seg03].

Metal Slivers

Metal slivers increased their presence with CMP (Chemical Mechanical Polishing). A small metal sliver lies between two interconnect lines barely, or not even touching them. These slivers can be from any of the metals used in IC an fab, such Al, Cu, W, or stainless steel. When temperature rises, metals expand, and the sliver now touches the signal lines. Higher voltages at burn-in can promote the rupture of the high resistance oxide surface of the metals, bonding the three metal elements [Rig98]. The bridge resistance is now permanent, and low enough to reduce noise margins or even cause functional failure.

Gate Oxide Shorts

Recent studies show that wearout and breakdown in the ultrathin oxides below 30 Å have different properties than for the thicker oxides on large channel length transistors [Deg01, Wei97, Seg95]. The older oxides were well characterized showing the elevation of I_{DDQ} and the reliability risk when gate shorts passed through the test process. Ultrathin oxides show a soft breakdown in addition to the hard breakdowns of thicker transistor oxides. Soft breakdown is an irreversible damage to the oxide whose most significant effect is the increase in noise of the gate voltage. I_{DDQ} is not elevated for the soft gate-substrate ruptures of ultrathin oxides. The noise can show up to four orders of magnitude increase after soft breakdown, and this is the only certain evidence of the irreversible damage to ultrathin oxides.

Degraeve, et al., found that uniformly stressed 2.4 nm gates oxides had a uniform area breakdown [Deg01]. Breakdown over the channel had a high resistance from $10^5 - 10^9$ **Ω**, while breakdowns over the drain and source had resistances of $10^2 - 10^4$. Since the drain and source gate overlap area was much smaller than the area over the channel, most of the breakdowns occurred over the channel. By measuring breakdowns in 150, 180, and 200 nm transistor, they found that the percentage of gate to drain and source breakdowns increased as the transistors became smaller. The drain and source gate overlap area becomes a larger fraction of the transistor gate area. The gate to drain or source breakdowns were likely to cause hard failure, while the gate to channel breakdown showed no failure effects. A 41-stage ring oscillator with seven transistor gate ruptures continued to function with a 15% decrease in oscillator frequency. Part of the frequency reduction was due to hot carrier damage during a pre-stress. Significantly, the ring oscillator

did not fail despite having several gate oxide breakdowns present.

Detection of softly ruptured ultrathin oxide does not appear possible at this time, nor is the reliability status clear. The normal functioning of the transistor with an ultrathin oxide is not as effected as were the killer ruptures of the thicker oxides. The recent ultrathin oxide experiments indicate test escape and subsequent reliabilities may not be as risky as for breakdown in older technologies [Deg01, Ala03, Rod03, Weir97. These different properties of the transistor oxide demand more studies at the circuit level to assess the implications of test escapes.

IV. Defect-Based Test (DBT) Approach for Nanotechnolgy ICs

DBT develops a detection strategy matching defect class electronic properties to particular test methods. The practice of DBT dates back in electronics history to when printed wire boards were visually inspected for solder blobs (bridges), wire trace opens, or mousebites. The tests were visual, but they targeted specific defects. The formal study of CMOS IC defect properties grew rapidly in the 1980's. A paper at International Test Conference in 1994 titled "Defect Classes – An Overdue Paradigm for Testing CMOS ICs" collated diverse data on bridges, opens, and parametric delay defects, and then proposed a unified test strategy based on these defect electronic properties [Haw94]. About the same time, Manual d'Abreu is believed to have first coined the phrase "Defect-Based Testing" at an Intel Corp. test workshop.

DBT differs from fault model approaches. A fault model hypothesizes how a circuit will fail and applies test patterns looking for a predefined failure. Defect based testing identifies the spectrum of defects that cause IC failure, and then develops test methods that are matched to the electronic properties of that particular defect class. An analysis for parametric failures was given in [Seg02].

The major weakness of traditional tests lies in setting a test limit when the parameter under measurement varies widely for intrinsic parts. Propagation delay, or other forms of speed testing, and I_{DDQ} values vary significantly for normal die so that the worst case intrinsic value as a test limit misses many defective die within this limit. This has driven post test analysis techniques. The major post test analysis developments have been with I_{DDQ} testing, $MinV_{DD}$ testing, and 2-temperature speed testing [Max99, Daa01, Mad02, Kes00].

I_{DDQ} has three testing approaches that are presently used in production tests. The delta-I_{DDQ} test simply subtracts a vector I_{DDQ} reading from its previous vector value. This tends to reduce background noise [Kru02, Mil00]. Agilent Corp. reported an I_{DDQ} production variance reduction scheme called current ratios [Max99]. The technique uses the I_{DDQ} vector distribution for many measurements per part. These distribution properties are retained as the chip average intrinsic background current varies considerably from die-to-die. A third production test technique was reported by a team from LSI Logic and Portland State University [Daa01]. The method uses the properties of nearest neighbor die to evaluate test limits for I_{DDQ}. This post-test processing technique is very sensitive. It has also correlated column and row data to further sensitize the test. The method is general, and it was also reported using $MinV_{DD}$ as a post test parameter [Mad02]. $MinV_{DD}$ is a measurement of the minimum V_{DD} that the IC will function at a given clock frequency. Outlier detection is obvious for many parts.

Another related test approach to identify outliers relies on the property that defective ICs don't track with the main population when speed (e.g., F_{MAX}, propagation delay) or I_{DDQ} measurements are taken for different temperature or V_{DD} values [Kes00]. This multi-parameter test approach is used most successfully in the detection of resistive vias.

V. Conclusion

The new nanotechnology CMOS ICs have old and new failure mechanisms. The newer ones can be difficult and expensive to detect. Modern testing is distinguished by intrinsic parametric variation, defects sensitive to detection only by a comparing a speed test at two or more operating parameters, or seeking outlier exposure in other ways such as the $MinV_{DD}$ test. Defect-based testing does not direct how to use a test in the presence of noise, but does tell you where to apply your multiparameter and statistical analysis resources. To this objective, we see a trend toward test techniques that identify outlier parts against the main population. This brings statistical analysis to a more significant role than in the past.

References

[Ala03] M. Alam and K. Smith, "A phenomenological theory of correlated multiple soft-breakdown events in ultr-thin gate dielectrics," *International Reliability Physics Symposium (IRPS),* pp. 406-411, April 2003.

[Bak99] K. Baker, G. Grounthoud, M. Lousberg, I. Schanstra and C. Hawkins, “Defect-based delay testing of resistive vias-contacts – a critical evaluation,” *International Test Conference (ITC),* pp. 467-476, October 1999.

[Ber98] K. Bernstein, K. Carrig, C. Durham, P. Hansen, D. Hogenmiller, E. Nowak, and N. Rohrer, *High Speed CMOS Design Styles*, Kluwer Academic Publishers, 1998.

[Bre96] M. Breuer and S. Gupta, “Process aggravated noise (PAN): New validation and test problems,” *International Test Conference (ITC),* pp. 914-923, October 1996.

[Bru02] M. Bruce, V. Bruce, D. Epps, J. Wilcox, E. Cole, P. Tangyungyong and C. Hawkins, “Soft Defect Localization (SDL) on ICs,” International Symposium on Test and Failure Analysis (ISTFA), November 2002.

[Daa01] R. Daasch, K. Cota J. McNames, and R. Madge, “Neighbor selection for variance reduction in I_{DDQ} and other parametric data,” *Int. Test Conf. (ITC),* pp. 92-100, Oct. 2001.

[Deg01] R. Degraeve, B. Kaczer, A. De Keersgieter and G. Groeseneken, “Relation between breakdown mode and breakdown location in short channel NMOSFETs,” *International Reliability Physics Symposium (IRPS),* pp. 360-366, May 2001.

[Deu01] A. Deutsch, et al., “On-chip wiring design challenges for gigahertz operation,” *Proceeding of the IEEE*, Vol. 89, No. 4, April 2001.

[ITR01] International Technology Roadmap for Semiconductors 2001 (http://public.itrs.net/).

[Kan01] K. Kana, K. Nose and T. Sakurai, “Design impact of positive temperature dependence on drain current in sub-1V CMOS VLSIs,” *IEEE Journal of Solid State Circuits,* pp. 1559-1564, October 2001.

[Haw94] C. Hawkins, J. Soden, A. Righter and J. Ferguson, “Defect classes – An overdue paradigm for testing CMOS ICs,” *International Test Conference (ITC)*, pp. 413-424, October 1994.

[Kes00] A. Keshavarzi, et al., “Multiple-parameter CMOS IC testing with increased sensitivity for I_{DDQ},” *Int. Test Conf. (ITC),* pp. 1051-1059, Oct. 2000.

[Kes02] A. Keshavarzi, K. Roy, J. Tschanz, S. Narendra, C. Hawkins, R. Daasch, M. Sachdev and V. De,” Leakage and Process Variation in Current-Based Testing on Future CMOS Circuits,” *IEEE Design & Test of Computers*, pp. 36-43, Fall 2002.

[Kru02] B. Krusman, R. van Veen and K. van Kaam, “The future of delta IDDQ testing,” ,” *Int. Test Conf. (ITC),* pp. 101-110, Oct. 2002.

[Mad02] R. Madge, et al., “Screening MinVDD Outliers Using Feed-Forward Voltage Testing,” *Int. Test Conf. (ITC),* pp. 673-682, Oct. 2002.

[Max99] P. Maxwell, P. O’Neill, R. Aitken, R. Dudley, N. Jaarsma, M. Quach, and D. Wiseman, “Current ratios: A self-scaling technique for production I_{DDQ} testing,” *Int. Test Conf. (ITC*), pp. 738-746, *October* 1999.

[Mil02] A. Miller, “I_{DDQ} testing in deep submicron integrated circuits,” *Int. Test Conf. (ITC*), pp. 724-729, *October* 1999.

[Nar99] S. Narendra, D. Antoniadis and V. De, “Impact of using adaptive body bias to compensate die-to-die Vt variation on within-die Vt variation,” *International Symposium on Low Power Electronics and Design*, pp. 229-232, 1999.

[Ors02] M. Orshansky, L. Millor, P. Chen, K. Keutzer and C. Hu, “Impact of spatial intrachip gate length variability on the performance of hi-speed digital circuits,” *IEEE Transactions on Computer-Aided Design,* Vol. 21, No. 5, May 2002.

[Rig98] W. Righter, C. Hawkins, J. Soden and P. Maxwell, “CMOS IC Reliability Indicators and Burn-In Economics,” *International Test Conference (ITC)*, Washington D.C., November 1998.

[Rod03] R. Rodriguez, J. Stathis and B. Linder, “Modeling and experimental verification of the effect of gate oxide breakdown on CMOS inverters,” *International*

Reliability Physics Symposium (IRPS), pp. 11-16, April 2003.

[Seg02] J. Segura, A. Keshavarzi, J. Soden and C. Hawkins, "Parametric failures in CMOS ICs – A defect-based analysis," *International Test Conference (ITC),* October 2002.

[Seg03] J. Segura and C. Hawkins, *CMOS Electronics: How it Works, How it Fails,* IEEE Press, Pub in Fall 2003.

[Sin95] A. Singh, H. Rasheed and W. Weber, "I_{DDQ} testing of CMOS opens: An experimental study," *International Test Conference (ITC),* pp. 479-489, October 1995.

[Syl01] D. Sylvester and K. Keutzer, "Impact of small process geometries on microarchitectures in systems on a chip," *Proceedings of the IEEE,* Vol. 89, No 4., pp. 467-489, April 2001.

[Van02] S. Vangal, et al., "5-GHz 32-bit integer execution core in 130-nm dual-V_T CMOS," *IEEE Journal of Solid-State* Circuits, vol.37, no.11, p.1421-32, Nov. 2002

[Wei97] B. Weir, et al., "Ultra-thin gate dielectrics: they break down, but do they fail?," *International Electron Device Meeting (IEDM),* pp. *73-76,* 1997.